Is there evidence for export-led adoption of ISO 14001? A review of the literature using meta-regression

Article (Accepted Version)

Is there Evidence for Export-Led Adoption of ISO 14001? A Review of the Literature Using Meta-Regression

Catherine Liston-Heyes[1] and Anthony Heyes[1]

1 University of Ottawa, Canada

Corresponding Author: Catherine Liston-Heyes, Graduate School of Public and International Affairs, Faculty of Social Sciences, University of Ottawa, 120 University Drive, Ottawa, Canada, K1N 6N5

Email: clistonh@uottawa.ca

Abstract

Does the export-orientation of a firm affect the likelihood that it adopts an environmental management certification? We use meta-regression methods to analyse systematically the corpus of published research on export-led adoption of the largest and most prominent certification, ISO 14001. We show that the explanatory variables authors' choose to include in their models reflect the tenets of stakeholder and institutional theories. We also find that the literature suffers from substantial publication bias but that, once this is accounted for appropriately, a genuine effect remains. The evidence from twenty years of published studies taken as a whole is that export do incentivise the adoption of the standard as often hypothesized by proponents of voluntary approaches and self-regulation.

Keywords

environmental management systems, export-orientation, ISO14001, meta-regression analysis (MRA), publication bias

The impact that globalization and increased international trade has on the natural environment and sustainability of business practices has been much debated in academic and practitioner circles (Kolk, 2016). Proponents of the view that trade enhances environmental performance appeal to a process whereby the regulatory standards and norms of highly regulated jurisdictions get transmitted to suppliers in less-regulated jurisdictions (Berliner & Prakash, 2014; Christmann & Taylor, 2001; Cole, Elliott, & Shimamoto, 2006; Prakash & Potoski, 2014). This is sometimes referred to as the "California effect" (Vogel, 1997).

As governance systems evolve, evidence suggests that this coercive effect may increasingly be transmitted by non-governmental and private actors through voluntary certification programs (Berliner & Prakash, 2015; Schembera, 2018). Participating firms voluntarily adopt higher standards than legally required, but in return program membership allows them to signal credibly their enhanced standards to stakeholders, for example socially-conscious consumers in the importing countries who would otherwise not be able to observe or evaluate internal practices (Prakash & Potoski, 2014). Voluntary accreditation programs are thus powerful policy instruments that facilitate the exchange of environmental stewardship for stakeholder appreciation (Vogel, 2005). The desire to export here incents a firm to want to certify it environmental practices.

The most prominent and widely-adopted international standard on environmental practices is ISO14001. To qualify requires that a firm demonstrate that it has in place an environmental management system (EMS) that satisfies a number of criteria, designed to ensure that environmental impact is managed in line with international good practice, and in a way that promotes continuous reflection and improvement. The standard is generic and does not have a single sectoral-focus. Depending upon where a firm primarily operates, ISO14001 accreditation may provide it with reputational and goodwill gains across the range of stakeholders including customers, suppliers, employees, local communities, NGOs, investors and regulators (Berliner & Prakash, 2015). In particular, ISO 14001 provides a credible signal of environmental stewardship to commercial audiences that transcends national borders (Berliner & Prakash, 2014; Christmann & Taylor, 2001; Prakash & Potoski, 2006). Accordingly, business scholars examining the determinants of ISO 14001 adoption have routinely included whether the firm "export" as an independent variable in regressions seeking to explaining ISO adoption. To date, there are 37 such studies involving a total 1 640 572 firms located around the world.

Our objective is to investigate the evidence base for the view that export-orientation acts as a driver of firm-level certification. A casual reading of this literature does not allow general insights to be drawn. Empirical studies vary in the hypotheses on which they choose to focus, the variables that are included as controls, and the ways in which variables are defined and measured. The challenge in drawing general lessons is further amplified by the diversity of datasets and methodological approaches that are used. We use established meta-analytic methods to investigate what the published literature tells us.

To sign-post our key findings, we find (1) the evidence base in this area is subject to significant positive publication bias, meaning that a naive reading of published papers would cause us to *over*-state the influence of export-orientation on adoption propensity. However, (2) evidence of a statistically significant positive *genuine* effect remains, even after accounting appropriately for that bias. In addition, (3) we find strong evidence that the quality of the methods used to estimate the relationship, the journal outlet where the study is published, and the publication date matter in terms of the size and significance of the effects. Better "quality" data, methods, and reviewing processes appear to facilitate the detection of a genuine export-led ISO14001 adoption effect.

The studies investigated use a diverse range of variables that might influence the ISO 14001 adoption decision of a firm. A breakdown of the determinants across the field reflects the influence of stakeholder theory contending that firms will respond to pressures exerted by stakeholders (Margolis & Walsh, 2003; Orlitzky, Schmidt, & Rynes, 2003; Prakash & Potoski, 2014; Schembera, 2018). These pressures can be conceived in a transactional perspective where firms make decisions on the basis of a comparison of the costs and benefits of certification or alternatively using neo-institutional theory according to which firms tend to adapt to prevailing norms and rules in its environment (Berliner & Prakash, 2015; Hauser & Hoenacker, 2014). Contributors to the literature sometimes omit detailed discussion of the theoretical foundations for the variables contained in the regressions that they report, though the compilation of variables suggests that both incentive and norm-based approaches are considered in this literature. The inclusion of institutional descriptors capturing the networks, industry sector, and the wider environment in which the firm operates are accounted for by many authors reflecting assumptions that external commercial audiences and domestic regulatory and stakeholder pressures influence certification in potentially different ways (Berliner & Prakash, 2014). Some implicitly recognize that exporting firms operate in both a home and at least one foreign country and are thus subject to different norms and rules that are internalized through mimetic, normative, and coercive processes (Hauser & Hogenacker, 2014; Kostova, 1999; Kostova & Zaheer, 1999). The behavior of internationally networked firms will consequently be affected by where and with whom they conduct their business and this appears to be of interest for some authors.

We start by outlining meta-regression in brief and describe the empirical literature that is the basis for this study. Next we detail how we identified and coded studies for inclusion in the statistical analysis and provide more details of the methods used in assessing publication bias and detecting a genuine effect. After we present and discuss the results, we present our conclusion.

Meta-Regression Analysis and the ISO 14001 Adoption Literature

This study is meta-analytic. Our objective is to formalise what can be inferred from the published literature in the area of export-orientation and environmental management certification.

Individual papers in this literature provide a review of research. Such a review is typically tailored to the focus of the particular paper, for example previous work that has explored a particular explanatory variable of interest, and implicitly adopts a "vote counting" approach (e.g., "five studies say this, two studies say the other") to give an impression of weight of evidence. Vote counting in the set of papers that we study would indicate that exports are a positive and significant determinant of the likelihood that a firm adopts ISO 14001. More concretely, 25 of 37 studies (68%) would support such a conclusion. Vote counting across the estimates that are indicated as significant by the authors yields a total of 88 of 141 (62.4%). Yet these papers differ considerably in the size of the sample they exploit, the provenance and age of the dataset, the modelling techniques they use and the factors that they are able (or choose) to control for. This means that such vote counting, though tempting, is inappropriate for reasons that are well-understood by meta-analysts (Combs et al., 2011).

Meta-regression analysis (MRA) uses multiple regression methods to identify publication bias and to explore whether evidence of a genuine effect survives once any such bias is taken into account. MRA is well suited to the synthesis of management research because the empirical work is rarely conducted in controlled experimental settings so there is considerable natural variation across studies. Moreover, publishing traditions in empirical management research are such that reporting protocols vary substantially between authors and journals. The ways in which researchers report and present data and how estimates were obtained are less standardised than in most natural sciences (Roberts 2005). This makes comparing results from two studies difficult. Combining results either implicitly when reading, or explicitly in the context of systematic review, is challenging (Higgins et al., 2003).

Publication bias refers to those elements of research practice, peer review, etc. that lead to the probability that a study is published to the size and statistical significance of the result. In general, larger and more statistically significant effects are over-represented in published literatures in many fields of inquiry. Brodeur, Sangnier, and Zylberberg (2016) provide compelling evidence of publication bias and "p-hacking" in top economics journals, while Harrison, Banks, and Pollack (2017, p. 400) conduct a topic-by-topic investigation of publication bias in the management literature and conclude that: "… publication bias affects many … topics in strategic management research. Correlation inflation due to publication bias ranged in magnitude from 0.00, indicating no bias, to 0.19, representing considerable bias".

Publication bias does not (necessarily) arise because authors are meaningfully "concealing" results. According to Card and Krueger (1995) reviewers, editors and authors are predisposed to treat results that are statistically significant, or generally expected, more favourably. In our setting, conventional wisdom is that the coefficient on exports will be positive and significant. Regressions not consistent with this may be excluded from a manuscript, may persuade the researcher to search for alternative empirical specifications that "work better," or be more likely to fail peer review. Stanley and Doucouliagos (2012) conjecture that publication selection bias may be unavoidable, making it important to correct for its adverse effects. Ignoring publication bias can distort literature review, whether is it a conventional narrative review or a meta-analysis (Higgins & Green, 2008). While it may be individually legitimate not to report counter-intuitive coefficients, the asymmetric selection of positive coefficients that ensues (the "file drawer problem") biases our collective understanding of the size of the true effect (Rosenthal, 1979).

MRA provides an objective approach to combine findings from heterogeneous studies to uncover the "… nuggets of truth that have settled to the bottom", in other words identify whether a genuine effect remains after publication bias has neutralized (Stanley & Doucouliagos, 2012, p. 3). An MRA practitioner works with "effect sizes" that measure the impact of variation in one variable on another holding other factors constant. These must be comparable between the studies that are being meta-analyzed (Becker & Wu, 2007). Studies of interest to us here are those that have the ISO 14001 certification decision of a firm as the dependent variable and a variety of explanatory variables on the right-hand side. These might include firm size, sector, export intensity and stakeholder metrics. However studies vary widely in which of these and other potential regressors they use and/or report, in addition to their statistical methods, the age and nature of dataset exploited, and so on.

As our interest is in the role of exports, the size effect pertinent for us is the coefficient associated with the export variable α_i in regression i. Given the heterogeneity in the way in which the coefficients are obtained across studies we instead use reported t-statistics associated with the export coefficients. The t-statistic is simply the coefficient divided by its standard error such that ($t_i = \frac{\alpha_i}{Se_i}$). It is a unitless measure of size effect comparable across studies. Such substitution is common practice and allows more studies to be retained. A caveat is that it is primarily a measure of statistical not "practical" significance of an effect. Nonetheless when used as a dependent variable in established meta-regression procedures it allows for identification of publication bias, detection of the existence of a genuine effect beyond bias and the uncovering of sources of heterogeneity of results between studies (Arestis, Chortareas, & Makonis, 2015; Bumann, Hermes, & Lensink, 2013; Card & Krueger, 1995; Mookerjee, 2006; Stanley & Doucouliagos, 2012).

While meta-analysis is long-established in medical and health sciences it is increasingly being used by management researchers (Carney et al., 2011; Hoobler et al., 2018; Orlistzky et al., 2003; Post & Byron, 2014). To the best of our knowledge, this is the first study using MRA to investigate the impact of exports on an environment-related variable.

Data and Methods

Data Collection

Stage 1 was the creation of a database of research articles ("the literature"). Two research assistants independently executed searches of the ECONLIT, ABI/INFORM, BUSINESS SOURCE COMPLETE and GOOGLE SCHOLAR (first 50 pages) looking for the terms "ISO14001," "14001," "ISO 14001," or "EMS" in item titles, abstracts and tags. They combined lists, discussing any non-overlap between them. Items recovered (at this stage over 600) included journal articles, working papers, conference proceedings and books. They read these and retained only empirical studies that contained at least one regression with a variable related to ISO 14001 as outcome variable. To avoid double counting overlapping working papers and conference proceeding versions of work later published as journal articles were removed. This reduced the pool of eligible studies to 71.

A further sort was carried out on dependent variable – the variable relating to ISO 14001 – and the export variable. Some papers do not examine the certification decision per se but rather the quality of the implementation or the extent of internalisation of the standard by firm managers. Others do not explicitly use exports as independent variable but a broader "consumer pressures" variable not intended to separate domestic from international buyers. These studies (n = 20) were excluded as they do not provide comparable size effects. Since our focus is firm-level accreditation we also discarded studies that aggregated certification across a jurisdiction (n = 12). Country-wide studies of ISO14001 are methodologically and theoretically different and use a different unit of analysis (Prakash & Potoski, 2006) that make them non-comparable with firm-level analysis. This left us with 37 firm-level studies that report regression results with ISO14001 adoption on the left-hand side and exports (among other variables) on the right.

In Stage 2 individual papers were coded following the methods used by Bellavance and colleagues (2009), Capon, Farley, and Hoenig (1990), Dalhuisen and colleagues (2003), Nelson and Kennedy (2009), Orlitzky and colleagues (2003), Wilson (2009), and others. The coding spreadsheet contained 82 variables. These included (1) document identifiers (authors, journal, year of publication), (2) characteristics of the data used (sample size, country), (3) characteristics of regression methodology (number of controls, estimation method, robustness checks), (4) details of the export variable and its coefficient and (5) measures of study quality (H-index of journal venue). In some cases, details were missing. When possible these were calculated retrospectively using information provided in the article (for example estimating the standard error from a reported t-statistic). In a small number of cases we contacted a study author for clarification. While use of t-statistics reduces the consequence of missing information we dropped two such studies due to incompleteness of data at the coding stage.

At Stage 3 the research team (two authors and two research assistants) met to review entries and resolve differences.

At the end of this process the database contained 141 estimates based on study of 1 640 572 firms derived from 37 studies. Appendix Table A1 reports these studies and key characteristics. These form the basis for the MRA that we conduct in this article.

With the database constructed we went on to investigate (1) whether the literature is subject to publication bias; (2) whether there is evidence of a genuine effect after accounting for any such bias; (3) whether variation in results between studies depend systematically on how the study was carried out and journal quality indicators.

Sample size plays a key role in collating results from different studies. Small samples typically yield estimates with higher standard errors (Se_i). Since $t_i = \frac{\alpha_i}{Se_i}$, a higher coefficient ($\alpha_i$) is needed to achieve any particular significance threshold (say 1.96). Larger samples may be expected to yield statistically significant estimates that are smaller in size as greater sample size delivers smaller standard errors. The detection of publication bias through these methods thus relies on the fact that "(W)hen publication selection is present, the reported effect is positively correlated with its standard error, ceteris paribus; otherwise, estimates and their standard error will be independent, as required by the conventional t-test and guaranteed by random sampling theory" (Stanley & Doucouliagos, 2012, p. 60).

Detecting and Adjusting for Publication Bias

The simplest and most commonly used method to examine publication bias is informal or "eyeball" examination of a funnel plot. A funnel plot is a scatter diagram of precision versus non-standardized size effects. There are alternative ways to measure precision though the most common and accurate is the inverse of the standard error ($1/Se_i$) associated with the coefficient of interest – in our case the coefficient on exports in the ISO 14001 adoption regression. Alternatives include the sample size (n), or its square root ($\sqrt{n}$) (Sterne & Egger, 2001; Sutton et al., 2000).

Absent publication bias, scatter points should shape into an inverted funnel as estimates vary randomly and symmetrically around the true population effect, regardless of its size (Sutton et al., 2000; Thompson & Higgins, 2002). Heteroscedasticity of errors imply that the plot spreads at the base as smaller studies typically have less precision. However, it is lack of symmetry which is indicative of publication bias. If selection is assumed to favour a certain direction, the funnel plot will have more observations to one side or the other.

While funnel plot examination is typically a compelling first step in looking for publication bias in a literature, meta-regression analysis (MRA) provides a complementary and more objective way to link effect size from a study to its associated standard error. More concretely, in the absence of publication selection, observed effects should vary symmetrically around the true value $\hat{B}_1$. Estimating the following equation on a population of studies,

$$Effect_i = \hat{B}_1 + \hat{B}_0 Se_i + \varepsilon_i \qquad (1)$$

generates the estimate $\hat{B}_0 Se_i$ which captures the relationship between the reported effects and the standard errors (i.e., the publication bias, Egger et al., 1997).

The problem with estimating equation (1) by ordinary least squares (OLS) is that it suffers from heteroscedasticity as errors vary between regressions.[1] The preferred approach to address this is to estimate it using weighted least squares (WLS) with an analytical weight of $1/Se_i^2$. This serves to correct for the estimated (heteroscedastic) variances.

An alternative is to divide Equation (1) by Se_i, which also constitute a sample estimate of the regression errors (Stanley, 2005). This yields:

$$t_i = \hat{B}_0 + \hat{B}_1(1/Se_i) + v_i \qquad (2)$$

where t_i are the t-statistics associated with the export regression coefficients. Note that the intercept and slope coefficients are reversed. Effectively, Equation (2) provides a statistical means of estimating funnel asymmetry. According to Egger and colleagues (1997), the estimate of the intercept $\hat{B}_0$ in this funnel-asymmetry test (FAT) provides a test of publication bias and its direction (Stanley, 2005). Rejecting statistically the null hypothesis that $\hat{B}_0$=0 is evidence in support of publication bias with the sign of the estimated coefficient indicating the direction of bias.

Note that because the independent variable $1/Se_i$ must be also be estimated, it is prudent to test additional specifications that replaces $1/Se_i$ with other measures of precision such as the square root of degrees of freedom ($\sqrt{Df}$) (Stanley, 2005). We also test $\sqrt{Df}$ as an instrument for $1/Se_i$ in a two stage least square (2SLS) alternative to Equation (2) (Davidson & MacKinnon, 2004).[2]

For the purposes of robustness, therefore, we use all of these procedures (funnel graphs, FAT with $1/Se_i$, FAT with $\sqrt{Df}$, and FAIVEHR) to investigate the existence, direction and extent of publication bias in our database.

We also probe heterogeneity by contrasting assessment of bias in subsample of studies that differ methodologically in terms of data, sample size, year of publication, journal quality, and attentiveness to robustness exercises.

If publication bias proves significant it can be filtered out from the existing literature by deflating each reported export coefficient by $\hat{B}_0 Se_i$.[3]

Detecting Genuine Effects

In additional to uncovering any publication bias we are also interested in determining the existence or otherwise of a true empirical effect of exports on ISO 14001. Again, for the purposes of robustness we use multiple approaches.

First, in Equation (2) we test whether we can reject the hypothesis H_1: $\hat{B}_1$=0 and in so doing confirm evidence favoring the existence of a genuine effect. This is referred to as precision-effect testing (PET, Stanley & Doucouliagos, 2012). More concretely, $\hat{B}_1$ provides us with the corrected effect, once publication bias has been filtered out or neutralized.

Second, we replace the t-statistics in Equation (2) by their corrected (for publication bias) version after suppressing the intercept term (i.e., by estimating the following regression)[4]:

$$corrected - t_i = \hat{B}_1(1/Se_i) + v_i \qquad (3)$$

If the estimate of $\hat{B}_1$ is statistically significant, we can conclude that a genuine effect exists. Again, we test this specification across a number of sub-samples to assess consistency of results across different strands of the literature.

Third, we use meta-significance testing (MST) to detect the presence of an overall empirical effect. MST uses the relationship between the t-statistics and the degrees of freedom (*Df*). We follow Card and Krueger (1995) by estimating the following equation:

$$E(Ln|t_i|) = \hat{b}_0 + \hat{b}_1 Ln(Df_i) + v_i \qquad (4)$$

This test is based on a property of statistical power which suggests that if H_1: $\hat{b}_1 \neq 0$, the log of the absolute value of the t-statistics will vary with the log of its degrees of freedom. More concretely, $\hat{b}_1=0$ when there is no genuine effect, and $\hat{b}_1=\frac{1}{2}$ when there is (Davidson & MacKinnon, 2004). Note that with publication bias $\hat{b}_1$ can take a value less than 0.5. However, if it remains greater than zero it implies a genuine effect. In other words, $\hat{b}_1>0$ implies a genuine effect.

Heterogeneity across Studies

Our final objective is to see what characteristics of a particular study might make it more or less prone to find an effect of exports on ISO adoption. Does the conclusion reached by a study depend systematically on how exports are computed or modelled, for example? Which variables it controls? For example, like any regression model, the estimates of MRA coefficients can become biased when important explanatory variables are omitted (Doucouliagos & Stanley, 2009; Efendic, Pugh, & Adnett, 2011). Consequently, we reduce the risk of omission bias in our MRA model by including moderator or control variables. Together, these variables capture factors that influence the magnitude of published effects and help explain the observed heterogeneity across studies (Bumann, Hermes, & Lensink, 2013; Mookerjee, 2006).

Following Arestis and colleagues (2015) we include dummy variables that account for the format of the export variable and data-related factors including study setting (geography and industry sector). We complement these methodological controls by including indicators that seek to control for study "quality." In particular, we monitor whether the results differ in studies that reported results of robustness checks versus those that did not and those that use estimation methods that rely on panel data. We also compiled four indices of journal quality where the study was published (ranking in SJR, AJG, ABS (2010) journal rankings, and the journal-level H-index) and test whether journal outlet affects the results.[5] In a similar vein, we examine whether studies published more recently (i.e., after 2009 – the midpoint in our dataset) differ from earlier attempts based on the notion that the data available on ISO14001 adoption grows through time (facilitating the use of panel data) and that later authors have the opportunity to learn from earlier work. Preliminary tests indicate that studies published after 2009 tend to appear in higher "rated" journals.[6]

In addition to data and study quality controls, we include a subset of apparently key explanatory variables used by authors to explain the variation in results. Selecting which to include is challenging. More than 50 different explanatory variables appear across the studies, sometimes just in one study. The mean number of controls per study is 20.

While the risk of omitted variable bias in the MRA suggests the inclusion of all relevant variables, with only 37 studies some specification searching is unavoidable (Higgins & Green, 2008). Appendix Table A2 lists the explanatory variables according to how frequently they are used in the database. We can see that many studies control for external stakeholder pressures (regulatory (58.9% of estimates), consumer (41.8%), community (29.8%), NGO (29.1%), investor (27.0%)), firm financial constraints (32.6%), firm and industry environmental performance/impact (32.6%, 29.1%) as well as foreign ownership/HQ (38.3%, 27.0%) as key explanatory factors. These author selections reflect theoretical assumptions and frames used in the wider literature including stakeholder theory (Margolis & Walsh, 2003; Orlitzky et al., 2003). Accordingly, we explore whether studies that account for regulatory, community, consumer, employee, NGO, and/or investor pressures deliver systematically different results. In the same vein, we tested whether studies that control for the CSR status (9.4%) of each firm deliver different results. The results from these analyses inevitably suffer from small sample problems, as we further slice-up an already quite small data set, so should be read only as suggestive.

Appendix Table A2 also indicates that most studies recognise that firms may be influenced by the context and institutions in the foreign countries with which they interact with and not only regulatory pressures at home (Berliner & Prakash 2014). They may adopt the behaviour they perceive as expected by the institutional environment in which they operate (Hauser & Hogenacker, 2014; Kostova, 1999; Kostova & Zaheer, 1999). Accordingly we created a portmanteau category called "international network" (IN) that takes the value of one if a study accounts for foreign activities other than exports (FDI, foreign ownership, foreign headquarters).

Data limitations inevitably hamper our ability to uncover omitted variable bias and the results from multivariate analysis are inconclusive. This is not simply a question of having a relatively small number of studies, but also the different ways in which studies vary. For instance, institutional theory suggests that normative, mimetic, and coercive pressures from within the industry can be powerful instruments of learning and compliance (Berliner &, Prakash 2015; Schembera, 2018). Accordingly, most authors (91%) include some form of industry control. However, these vary substantially in nature and number with some studies including 24 different controls. For this reason, many suppress them from their tables (Cole et al., 2006; Grolleau, Mzoughi, & Pekovic, 2015; King, Lenox, & Terlaak, 2005; McGuire, 2014; Nakamura, Takahashi, & Vertinsky, 2001; Nishitani, 2009). Those that include them, list their industry/economic sector in ways that do not overlap across studies. There is no real solution to this – our hands are tied by the way in which the underlying studies have been executed and reported.

We additionally consider if results are sensitive to whether a study focusses on the manufacturing sector (most common) or a broader set of firms, and (slightly different) whether a study includes firms that operate exclusively in the manufacturing sector (excluding retail and services sectors) versus those operating across such sectoral boundaries. Given that this is study-specific, the controls introduced to capture these elements are methodological.

Following Stanley and Doucouliagos (2012), Doucouliagos and Stanley (2009) and Arestis and colleagues (2015), we embed these k study descriptors in the FAT-PET MRA (Equation 2) as follows:

$$t_i = \hat{B}_0 + \hat{B}_1(1/Se_i) + \sum \hat{B}_k Z_{ki}/Se_i + v_i \qquad (5)$$

and into the MTS MRA (Equation 4) as follows:

$$E(Ln|t_i|) = \hat{b}_0 + \hat{b}_1 Ln(Df_i) + \sum \hat{b}_k Z_{ki} + v_i \qquad (6)$$

Variables are added in four different blocs in the nested multivariate regressions. Bloc A introduces the three measures of precision ($1/Se_i$, $\sqrt{d}f$, and *Ln(df)*), Bloc B contains data indicators, Bloc C adds key explanatory variables found across the literature, while Bloc D introduces study quality indicators. The added explanatory power of these blocs is assessed and discussed.

All studies report more than one regression. While some authors advocate taking a single representative regression per study, others recommend keeping the regressions as separate data points (Horvathova, 2010; Rathner, 2013). Nelson and Kennedy (2009) observe that the majority of meta-analyses in environmental economics include multiple observations per study. We do so here to avoid discarding information but correct for this by treating each study as a "cluster" and computing cluster-robust standard errors throughout the analyses.

Results

Descriptive statistics are presented in Appendix Table A3. Of the 141 estimates that we examine the t-statistics associated with the export coefficients have a mean of 2.68 and a standard deviation of 4.70. Since Study 34 (Appendix Table A1) is an important outlier with respect to its sample size (n=41 553 firms) and in its reported t-statistics, it is omitted from some specifications.

Figures 1 through 6 are scatter plots of precision measured by the (1) inverse of the standard error associated with the export coefficients, (2) sample size and (3) square root of sample size - against the estimated export coefficient. As explained, since the dependent and independent export variables differ across studies, direct comparison of export coefficients are likely to mislead.[7] Consequently, we also plot these precision measures against partial correlation coefficients (Stanley & Doucouliagos, 2012).[8]

In Figures 7 through 12 we split regressions drawn from studies subsampled in three different ways. First, those published in journals with high or low H-indices. Second, whether the study uses cross section or panel data. Third, whether the study was published before or after 2009 (the mid-point in our data series). We also generated plots using other study differentiators but all produced relatively similar shapes (i.e., not the symmetric funnel associated with a literature that is without publication bias). Eyeball examination of Figures 1 through 12 show elongated right tails suggestive of publication bias. Regressions delivering negative coefficients on the export variable are less likely to be published.

____________________Insert Figures 1-12 here_________________________

As discussed, the funnel asymmetry test (FAT) can bolster visual assessments. The estimates obtained from Equation 2 are reported in Table 1. The main result in this table is reported in the first column. That the intercept estimate is positive and significant indicates a statistically significant and positive publication bias, consistent with visual inspection of the funnel plots. The second column reports the same exercise but excludes the outlier (Study 34), with main result undisturbed, assuaging any concern that conclusion is being driven by that very large sample study.

The remaining columns in Table 1 report secondary results. The breaking up of the sample here means a number of the tests are under-powered, so we should interpret results – especially failure to find bias - cautiously. However results are consistent across the board. We see statistically significant evidence for publication bias in results drawn from studies using cross-sectional and panel methods, smaller and larger samples, published in lower and higher H-index journals, those with and without reported robustness checks and those published before and after 2009.[9] We believe our adopted approach to measuring precision is the most appropriate here. However, for completeness in Appendix Table A4 we replace $1/Se_i$ by $\sqrt{df}$ as an alternative precision measure (Stanley, 2005). This delivers mixed results. In most cases the estimated intercepts are either insignificant or positive and significant, consistent with upward publication bias. However, the correlation between $1/Se_i$ and $\sqrt{df}$ is very low (r = 4.0%), making $\sqrt{df}$ a poor proxy for $1/Se_i$.

____________________Insert Table 1 here__________________________

In summary we find strong evidence of upward publication bias in this literature leading it to significantly overstate the role of exports in firm-level ISO adoption decisions.[10]

To identify whether a genuine effect persists after adjustment for publication bias we examine the slope of meta-regression represented in Equation (2). These values are reported in the second row in Table 1. We see a significant (though only at 10%) positive genuine effect in the whole sample estimate reported in column 1, the significance of which is substantially improved by removal of the outlier study (second column). Looking at the subsample exercises in the other columns, we see that in all 10 cases coefficient estimates are positive, though in only 5 cases is that coefficient statistically significant. We are cautious not to over-interpret these underpowered tests, particularly where they fail to deliver significance. However, taken at face value they say that the larger and more significant estimates of genuine effect (after filtering for publication bias) come from studies using larger samples, reporting more robustness checks and published after 2009. Quality of journal (at least as captured by H-index) makes little discernible difference – we see significant genuine effects from papers in both halves. With the strong caveat noted in the last paragraph, we repeat this exercise using $\sqrt{df}$ as a precision proxy in Appendix Table A4. In this case we see that the estimated slope coefficient is positive in each of the 12 columns, but only achieves significance at conventional levels in the second column.

So evidence from this exercise is that after accounting for publication bias there remains an effect, albeit smaller than typically claimed in the literature. To further interrogate we additionally use corrected t-statistics – obtained after filtering the publication bias from the export coefficients – in a regression that is forced through the origin by suppressing the intercept term (Equation 3). The results of this are reported in Table 2. The estimated slope coefficient is now positive in each of the 12 columns and significant in all but one. This point strongly towards a genuine effect remaining, after the stripping out of publication bias, with that insight again coming from virtually all of the various sub-categorised studies.

___________________Insert Table 2 here________________________

Our final way of testing for genuine effect is through meta-significance testing (MST) which exploits the relationship between the log of the absolute value of t-statistics and the log of degrees of freedom (*df*). These results are reported in Table 3. Of interest to us here are the terms in the second row which are the coefficients on the *Ln(df)* variable. Note that in both of the main specifications – the first two columns – the genuine effect is confirmed. Looking across the subsample analyses we have positive estimated coefficients and significance, often at a very high level, in most columns. In one case (large-sample studies) we observe a negative coefficient, though small in absolute value and far from significance at conventional levels. Taken in the whole we find statistically significant estimates fluctuate between 0.54 and 0.20. According to Davidson and MacKinnon (2004) $\hat{b}_1$=.50 is an indication of a genuine effect but in the presence of publication bias, this estimate will be lower but remain above zero.

__________________Insert Table 3 here_______________________

Read as a group, Tables 1-3 point to studies using cross sectional methods ascribing weaker significance to a genuine effect. On the other hand, studies that use panel data are more likely to uncover a genuine effect. Panel methods allow the researcher to exploit within-firm variation that are lost in cross-section and identify endogeneity biases. For instance, researchers in international business suggests that firms doing business abroad face a "liability of foreignness" (i.e., costs arising from the unfamiliarity of the environment and the need for coordination across geographic distances among other factors, Zaheer, 1995). To overcome this disadvantage and compete successfully against local firms, exporting firms will require additional resources and/or stronger organizational or managerial capabilities (Barney, 1991) which may be correlated with unobservable factors encouraging ISO 14001 adoption. Hence, cross section studies may be more prone to endogeneity biases compared to studies that use panel data that allow for meaningful use of firm fixed effects. Moreover, panel data allows researchers to differentiate between early and late adopters (Baek, 2017), tracking intertemporal changes in motivations and in environmental factors such as fines and regulations (Blackman & Guerrero, 2012). That studies published in higher H-index journals are more likely to uncover a genuine effect, perhaps encouraged by editors and reviewers to use more sophisticated methods, is also reassuring of the view that such an effect does indeed exist.[11]

Thus far we have provided what we believe to be quite convincing evidence that (1) there exists publication bias favouring results that show a positive and significant role for exports in the accreditation decisions of firms and, (2) that after adjusting appropriately for that bias there remains good evidence of a genuine effect, though smaller than an averaging of claims in the literature would suggest. Moreover, (3) we show that a genuine effect is more easily identified in papers using better methods and/or published in better journals.

Our final analysis probes the question of why studies differ in their findings. In doing so, we implicitly recognise that estimates can vary for other reasons than publication bias, for example as a function of methods, model selection, data and timing (Jarrell & Stanley, 2004).

As almost always the case in empirical work the estimates presented may be biased if explanatory variables correlated with exports are omitted (Efendic et al., 2011). The export coefficient in individual regressions may erroneously be credited by a researcher for the influence of correlated omitted variables. In our database, there is a statistically significant negative correlation (r = -.247**) between the number of controls contained in a regression and the value of the coefficient derived. While crude this suggests that adding more controls tends to mitigate the size of effect attributed to exports.

Accordingly, the next objective of this article is to embed the MRAs into a more fully specified nested multivariate model. We do so using amended versions of the FAT-PET (Equation 2) and the MTS (Equation 4) discussed above. We proceed by adding the data controls to the precision indicators (Bloc B), followed by the explanatory variables (Bloc C), and study quality variables in a fourth step (Bloc D). The nested models are estimated using the alternative measures of precision. Finally, we present parsimonious versions of the extended regressions, where variables that have not been significant in any of the specifications are discarded (Arestis et al., 2015).

Table 4 reports the results of the nested multivariate regression analyses using cluster robust errors. Firstly, we note that the intercept $\hat{B}_0$ in the FAT-PET regressions remains positive and significant across all specifications, consistent with the existence of publication bias in the literature.

The coefficient $\hat{B}_1$ – which tests whether a genuine effect exists once publication bias has been neutralized – remains positive in almost all of the specifications, including in the three parsimonious/preferred versions (Models E.I, E.II, and E.III). This is important since the inclusion of additional variables can "dilute" the publication and genuine effects.

While investigating publication bias and existence of a genuine effect are usually the primary motivation for such multivariate regressions, the approach provides additional insights into the characteristics of a study that make the authors more or less likely to infer a positive impact of exports on accreditation.

Firstly, we note that the incremental explanatory value of variables organised by blocs (data indicators – Bloc B, regressors – Bloc C, and quality indicators – Bloc D) is statistically significant throughout the formulations except for Model BIII. In this particular case, adding the three data indicators does not significantly improve the explanatory power of the model. Of particular interest, we highlight the contribution of the quality indicators Bloc D on the overall explanatory performance of the models (Models DI, DII, DIII).

Examining the contents of each blocs, we notice that the variables from Bloc B are rarely significant as we moved towards the fuller specifications in Table 4. We note however that export coefficients extracted from datasets of Asian firms are associated with lower values of t. This is surprising since the penetration of ISO certification is particularly high in many key Asian countries, and firms in those countries are striving to grow exports to Europe and North America, where certification is often valued (Christmann & Taylor, 2001). However, further analyses conducted on this subset of studies (n=21) show that they typically involve significantly smaller datasets - which decreases the size of t-statistics relative to other studies – although the detection of publication and a genuine effect remains throughout the different approaches (MRA and MST tests).

With regards to the individual indicators that capture the variables and controls included in various studies (Bloc C), the signs must also be interpreted with great caution. The indicators for CSR, economic performance, industry environmental impact, investor, and community pressures collectively capture the environment in which firms operate and will vary with each data point within a particular study. Nonetheless, taken at face value, the results in Table 4 suggest that, with the exception of community pressure, excluding any of these measures from the regressions will generally decrease the t-ratios associated with the export variables. Poor or absent CSR leadership, investor pressures, and/or financial constraints may hamper the certification process and these effects impact upon the export coefficients when they are omitted. We note that there is a wide set of explanatory variables used by the various authors in different combinations, and the literature (therefore database) is not sufficiently large to allow the role of particular controls to be isolated nor is it possible to test every potential combination.

As for the study quality indicators contained in Bloc D, we note that studies published in higher H-index journals are associated with lower t-ratios. We already established above that studies with high H-index consistently indicate genuine effect. The negative sign here is not necessarily in conflict with this finding. Recall that t-ratios are simply the coefficient on the export term divided by its standard errors ($t_i=\alpha_i/Se_i$) and hence they will decrease either through lower coefficients and/or higher Se_i. Studies published in higher quality journals may be less prone to problems associated with omitted variables and subject to more rigorous testing that disentangles the effects of exports on adoption from other factors. Second, studies published after 2009 are associated with higher t-ratios. This potentially reflects the accumulation of data through time (recall that ISO 14001 was introduced in 1996) and its associated increase in sample size and opportunity for panel data. It can also reflect the changing behavior of firms as the certification gains popularity and/or the evolution of empirical models. Third, with respects to methods we find that coefficients from simple probit and logit methods are less favorable to exports. These studies typically use a dichotomous ISO adoption variable (the most common specification) as opposed to a continuous version that measures levels of completion, intentions, or stated probability of adoption within a given time period. They are also more likely to use cross section data (as verified by cross-tabulations). Hence, papers using such methods are less likely to dissociate true export effects on certification from others. This could explain the negative and significant sign observed through almost all the specifications that include this descriptor.[12]

It is important to note that reported t-statistic associated with the coefficient on the export variable as our dependent variable in meta-regressions has the advantage of being comparable across estimates which allows for larger sample sizes. However, it is a purely statistical measure and does not capture the economic significance of the estimated coefficient. More concretely, while our findings provide valuable insights on the scale and nature of statistical heterogeneity across adoption studies, they do not allow conclusions to be drawn on the economic importance of these effects.

__________________Insert Table 4 here_______________________

Conclusions

Certification of environmental practices is widely regarded as encouraging sustainability and better environmental practices, particularly by firms operating in countries where regulatory standards are comparatively lax and/or under-enforced. An important question, then, is how far export-orientation and the internationalisation of supply chains will propagate higher environmental standards.

One of the ways through which environmental"best practices" transfer from place to place is through the diffusion of ISO 14001 – the most widely adopted certification scheme for environmental management systems (Prakash & Potoski, 2006). This voluntary accreditation scheme is a powerful policy instrument that helps overcome problems of asymmetric information that arise in buying goods from an unfamiliar or less stringently regulated jurisdiction allowing consumers in the importing countries to impose their desire for "green" sourcing on foreign producers (Darnall, 2006; King et al., 2005), a variation of the California Effect (Vogel, 1997, 2005).

The intensity of stakeholder demands combined with the institutions and networks that propagate the norms that frames firms' environment ultimately determines the appeal of the certification (Berliner & Prakash, 2014, 2015; Hauser & Hoenacker, 2014; Prakash & Potoski, 2014; Schembera 2018, Orlitzky and Schmidt 2003, Margolis and Walsh 2003). Accordingly, dozens of empirical studies have been published using exports as a key decision factor in the certification decision.

So does the export orientation of a firm actually influence the likelihood that it adopts ISO certification of its environmental practices? In this study we reviewed systematically the available evidence – 141 regressions run on data relating to 1 640 572 firms and published in 37 studies. Using well-established methods for meta-analysis we show that: (1) the evidence base in this area is subject to significant positive publication bias, meaning that a naive reading of published papers would cause us to *over*-state the influence of export-orientation on adoption propensity. However, (2) evidence of a statistically significant positive *genuine* effect remains, even after accounting appropriately for that bias. (3) We also find that studies that use panel data and related methods that account for fixed effects are more likely to find a genuine effect as are studies published in better journals (as proxied by the H-Index) and studies published more recently. These are strong indications that the peer-reviewed publication process appears to be working with respect to this literature.

We tested whether the authors' choice of exogenous regression variables has an impact on the regression results. We concluded that community, investor, financial, and industry environmental impact as firm CSR status were useful in explaining variations in the strength of the relationship between exports and certification. However, data limitations mean that results from this element of the article need to be treated with caution, as small sub-sample sizes implied under-powered tests for individual factors.

As in Margolis and Walsh (2003) in relation to the CSP-CFP link, we end by highlighting potential unintended consequences of inferences based on the empirical validation of the export-

ISO14001 adoption relationship. In particular, it may falsely suggest that exports lead to healthier domestic environments, that firms who do not exports, pollute more, and/or that market-based solutions "work." Our findings do not establish an empirical relationship between the export behaviour of firms and their environmental performance.

Notes

1. Heteroscedasticity occurs when regression errors vary with the effects being modeled, thereby biasing the statistical significance of the results.

2. Stanley (2005) refers to these heteroscedasticity-robust funnel asymmetry instrumental variables estimators as FAIVEHR.

3. Note that the correction term B ̂_0 〖Se〗_i is calculated using absolute values of the effects and associated t-values.

4. The regression is forced through the origin since systematic bias has, in principle, been removed.

5. Few things in the academic profession are as contentious as journal rankings. We take no position on their merits here, but report these as commonly cited such rankings. The SCImago Journal & Country Rank (SJR) is a publicly available portal that includes the journals and country scientific indicators developed from the information contained in the Scopus® database (Elsevier B.V.- https://www.scimagojr.com/journalrank.php). The AJG is produced by the Chartered Association of Business Schools (UK) as is the ABS (2010) list. The H-index was obtained from the SJR website.

6. In effect, we tested a much wider range of method and data controls including controls for endogeneity, number of independent variables, various geographical indicators of data origins (country, region, continent, level of development, OECD), average age of the data, period spanned by the data, number of countries the data was sourced from. The indicators retained above carried the greatest explanatory power.

7. In our sample, 48% of regressions use a dichotomous ISO 14001 certified – not certified dependent variable and 43% use a dichotomous export – no export variable. The overlap between the two sets is only 18%.

8. Partial correlation coefficients are rarely reported in management studies but can be computed from reported regression statistics using the following equation:

9. $r=t/\sqrt{t^2+df}$

10. where t denotes the t-statistic and df, the degrees of freedom associated with the export coefficient.

11. Aside from the cross section/panel data and robustness checks (yes/no) sub-samples, the dividing benchmarks (study sample size, H-index, and published) were selected to equalize the number of studies in each category.

12. Egger and colleagues (1997) recommends the use of less-demanding significance levels. Accordingly we demarked coefficients that are significant at the $p<.10$ with an “x.”

13. We are grateful to an anonymous referee for bringing this to our attention.

14. The estimation method descriptor conflicted with the cross vs panel data indicator. As it performed better across a wider set of procedures, it was retained as our preferred variable.

Acknowledgements

We are particularly grateful to Dwayne Schindler, Steve Martin, Ariadna Cornei, Philippe Dorion, Rick French, the editor, and the referees of this journal for very constructive comments. Errors are ours.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Acknowledgement

The authors received no financial support for the research, authorship, and/or publication of this article.

References (see Appendix Table A1 for MRA articles)

Arestis, P., Chortareas, G., & Magkonis, G. (2015). The financial development and growth nexus: A Meta-analysis. *Journal of Economic Surveys*, *29*, 549-565.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*, 99-120.

Becker, B. J., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, *22*, 414-429.

Bellavance, F., Dionne, G., & Lebeau, M. (2009). The value of a statistical life: a meta-analysis with a mixed effects regression model. *Journal of Health Economics*, *28*, 444-464.

Berliner, D., & Prakash, A. (2014). Public authority and private rules: How domestic regulatory institutions shape the adoption of global private regimes. *International Studies Quarterly*, *58*, 793-803.

Berliner, D., & Prakash, A. (2015). "Bluewashing" the firm? Voluntary regulations, program design and member compliance with the United Nations Global Compact. *Policy Studies Journal*, *43*, 115-138.

Brodeur, A., M. Le, M. Sangnier, & Y. Zylberberg (2016). Star Wars: The empirics strike back. *American Economic Journal: Applied Economics*, *8*, 1 - 32.

Bumann, S., Hermes, N., & Lensink, R. (2013). Financial liberalization and economic growth: A meta-analysis. *Journal of International Money and Finance*, *33*, 255-281.

Capon, N., Farley, J. U., & Hoenig, S. (1990). Determinants of financial performance: a meta-analysis. *Management Science*, *36*, 1143-1159.

Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: a meta-analysis. *The American Economic Review*, *85*, 238-243.

Carney, M., Gedajlovic, E. R., Heugens, P. P., Van Essen, M., & Van Oosterhout, J. H. (2011). Business group affiliation, performance, context, and strategy: A meta-analysis. *Academy of Management Journal*, *54*, 437-460.

Cole, M. A., Elliott, R. J., & Shimamoto, K. (2006). Globalization, firm-level characteristics and environmental management: A study of Japan. *Ecological Economics*, *59*, 312-323.

Combs, J. G., Ketchen Jr, D. J., Crook, T. R., & Roth, P. L. (2011). Assessing cumulative evidence within 'macro' research: Why meta-analysis should be preferred over vote counting. *Journal of Management Studies*, *48*, 178-197.

Dalhuisen, J. M., Florax, R. J., De Groot, H. L., & Nijkamp, P. (2003). Price and income elasticities of residential water demand: a meta-analysis. *Land Economics*, *79*, 292-308.

Darnall, N. (2006). Why firms mandate ISO 14001 certification. *Business & Society*, *45*, 354-381.

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. Oxford: Oxford University Press.

Doucouliagos, H., & Stanley, T. D. (2009). Publication selection bias in minimum wage research? A meta-regression analysis. *British Journal of Industrial Relations*, *47*, 406-428.

Efendic, A., Pugh, G., & Adnett, N. (2011). Institutions and economic performance: A meta-regression analysis. *European Journal of Political Economy*, *27*, 586-599.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629-634.

Harrison, J., Banks, G.C., & Pollack, J. (2017). Publication bias in strategic management research. *Journal of Management*, *43*, 400-425.

Hauser, C., & Hogenacker J. (2014) Do firms proactively take measures to prevent corruption in their international operations? *European Management Review 11*, 223-37.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557.

Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions* (Volume 5). Chichester, England: Wiley-Blackwell.

Hoobler, J. M., Masterson, C. R., Nkomo, S. M., and Michel, E. J. (2018). The business case for women leaders: Meta-analysis, research critique and path forward. *Journal of Management*, *44*, 2473-2499.

Horváthová, E. (2010). Does environmental performance affect financial performance? A meta-analysis. *Ecological Economics*, *70*, 52-59.

Jarrell, S. B., & Stanley, T. D. (2004). Declining bias and gender wage discrimination? A meta-regression analysis. *Journal of Human Resources*, *38*, 828-838.

Kolk, A. (2016). The social responsibility of international business: From ethics and the environment to CSR and sustainable development. *Journal of World Business*, *51*, 23-34.

Kostova, T. (1999). Transnational transfer of strategic organizational practices: A contextual perspective. *Academy of Management Review*, *24*, 308-324.

Kostova, T., & Zaheer, S. (1999). Organizational legitimacy under conditions of complexity: The case of the multinational enterprise. *Academy of Management Review*, *24*, 64-81.

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly*, *48*, 268-305.

Mookerjee, R. (2006). A meta-analysis of the export growth hypothesis. *Economics Letters*, *91*, 395-401.

Nelson, J. P., & Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment. *Environmental and Resource Economics*, *42*, 345-377.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, *24*, 403-441.

Post, C., & Byron, K. (2015). Women on boards and firm financial performance: A meta-analysis. *Academy of Management Journal*, *58*, 1546-1571.

Prakash, A., & Potoski, M. (2006). Racing to the bottom? Trade, environmental governance, and ISO 14001. *American Journal of Political Science*, *50*, 350-364.

Prakash, A., & Potoski, M. (2014). Global private regimes, domestic public law: ISO 14001 and pollution reduction. *Comparative Political Studies*, *47*, 369-394.

Rathner, S. (2013). The influence of primary study characteristics on the performance differential between socially responsible and conventional investment funds: A meta-analysis. *Journal of Business Ethics*, *118*, 349-363.

Roberts, C. J. (2005). Issues in meta-regression analysis: An overview. *Journal of Economic Surveys*, *19*, 295-298.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638.

Schembera, S. (2018). Implementing corporate social responsibility: Empirical insights on the impact of the UN Global Compact on its business participants. *Business & Society*, *57*, 783-825.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*, 309-345.

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-Regression Analysis in Economics and Business* (Vol. 5). Routledge. p.186.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*, 1046-1055.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester; New York: J. Wiley.

Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*, 1559-1573.

Vogel, D. (1997). Trading up and governing across: transnational governance and environmental protection. *Journal of European Public Policy*, *4*, 556-571.

Vogel, D. (2005). *The market for virtue*. Washington, DC: Brookings Institution Press.

Wilson, D. B. (2009). Systematic coding. *The handbook of research synthesis and meta-analysis*, Russell Sage Foundation, 2, 159-176.

Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management Journal*, *38*, 341-363.

Author Biographies

Catherine Liston-Heyes (PhD, McGill University) is Professor at the Graduate School of Public and International Affairs at the University of Ottawa. From 1993 to 2011, she has been a member of faculty at the School of Management at Royal Holloway College, University of London. Her research interest focus on corporate social responsibility, regulatory economics, and the environment. Her articles have appeared in many journals including *Business Strategy and the Environment, Journal of Business Ethics, Journal of Environmental Management, Journal of Public Economics, Journal of Regulatory Economics, NonProfit and Voluntary Sector Quarterly,* and *Public Management Review.*

Anthony Heyes (PhD, McGill University) is Professor of Economics and Canada Research Chair in Environmental Economics at the University of Ottawa and part-time Professor of Economics at the University of Sussex. He has worked on corporate social responsibility, environmental regulation, and other topics. He has published in many journals including *American Economic Review, Journal of Health Economics, Journal of Public Economics,* and *Management Science*.

Fig. 1: Funnel Plot (Export Coefficients, 1/SE)

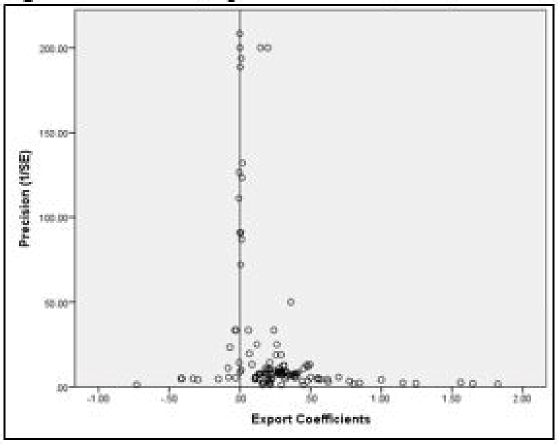

Fig. 2: Funnel Plot (Partial Correlations, 1/SE)

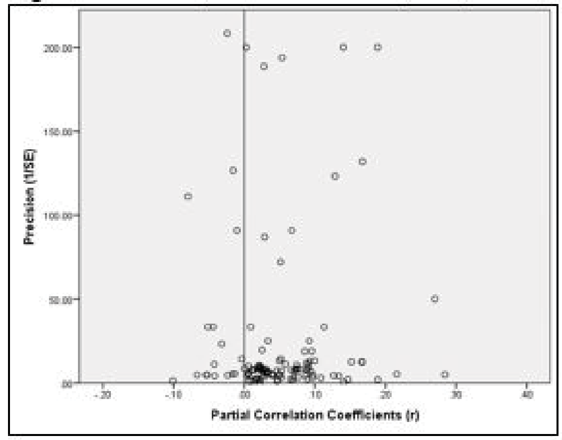

Fig. 3: Funnel Plot (Export Coefficients, Sample Size)

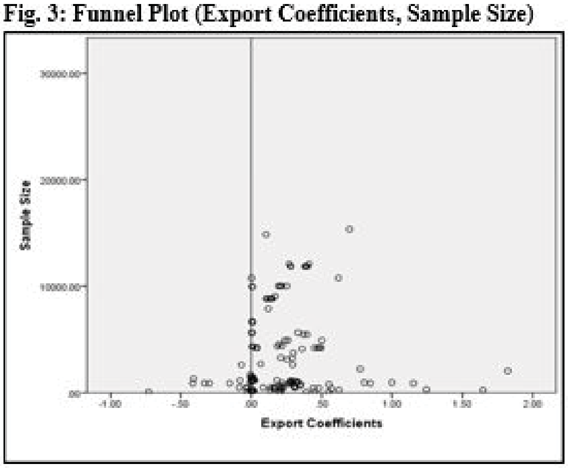

Fig. 4: Funnel Plot (Partial Correlations, Sample Size)

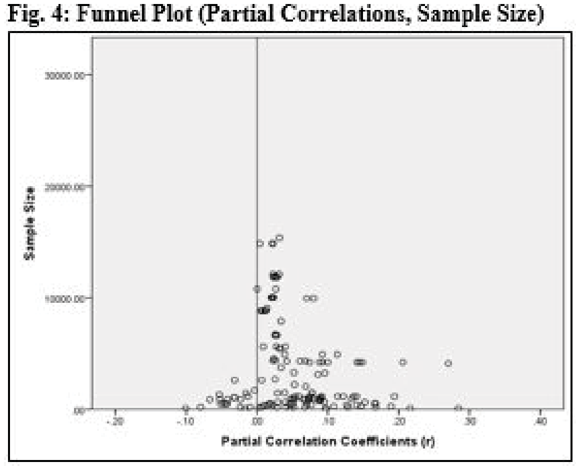

Fig. 5: Funnel Plot (Export Coefficients, $\sqrt{n}$)

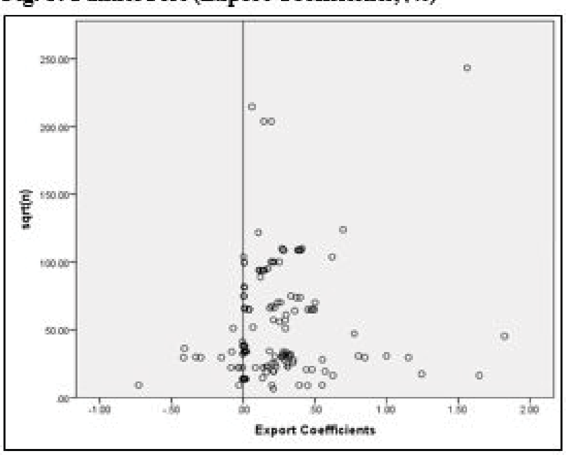

Fig. 6: Funnel Plot (Partial Correlations, $\sqrt{n}$))

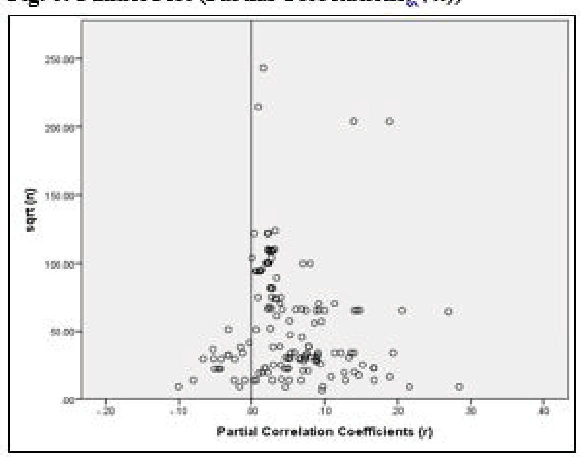

Fig. 7: Funnel Plot (Journal H-index above median)

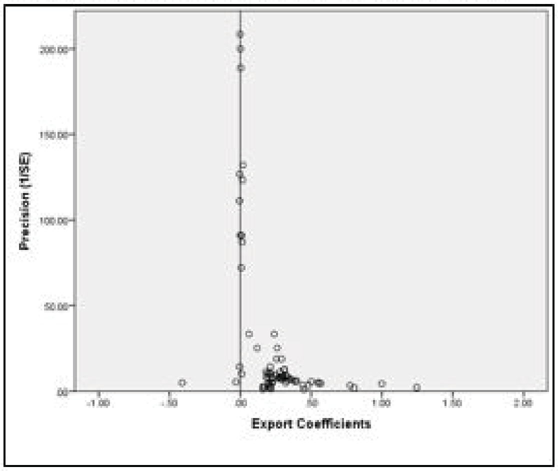

Fig. 8: Funnel Plot (Journal H-index below median)

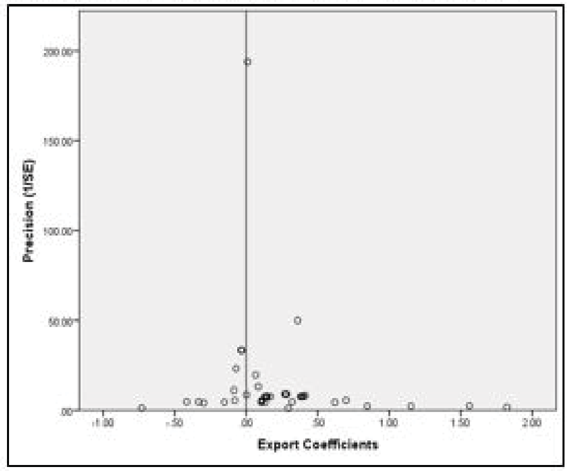

Fig. 9: Funnel Plot (Cross Section Data)

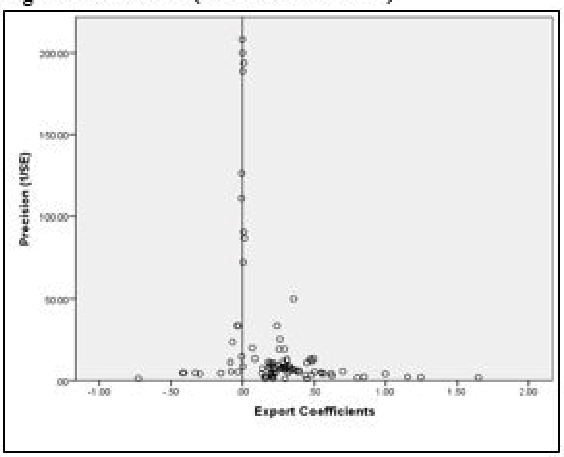

Fig. 10: Funnel Plot (Panel Data)

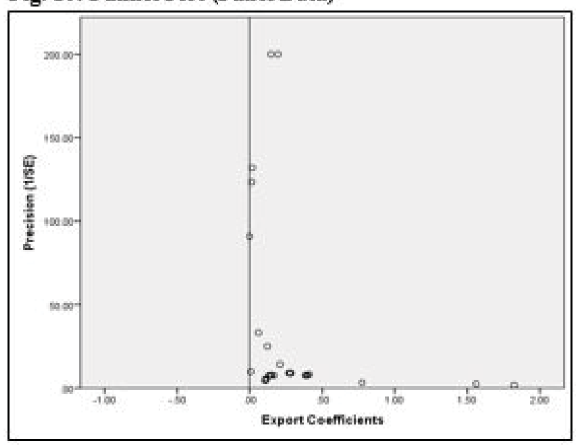

Fig. 11: Funnel Plot (Published before 2009)

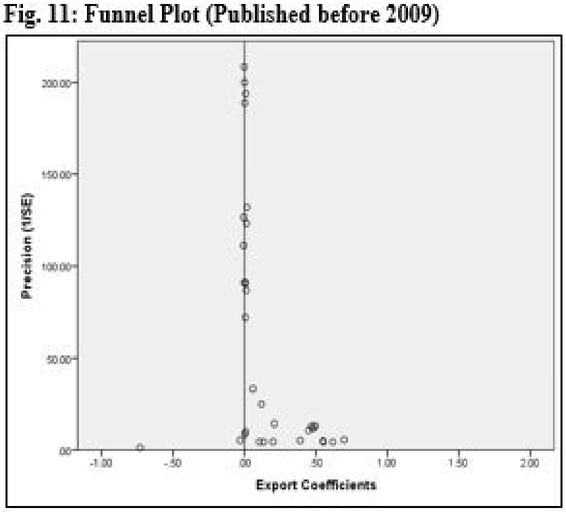

Fig. 12: Funnel Plot (Published since 2009)

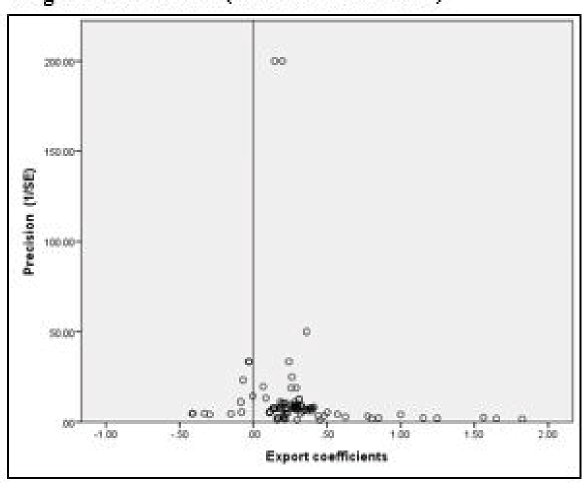

Table 1: MRA (FAT and PET) tests for publication selection (dep. var. t-stats)[a,b,c]

	All studies	Without outliers	Cross section	Panel data	Smaller samples	Larger samples	Low HIndex	High HIndex	Robust checks	No checks	Before 2009	After 2009
Intercept ($\hat{B}_0$)	2.2438*** (.5236)	1.8988*** (.3799)	1.9719*** (.5242)	1.7563*** (.2184)	1.1837*** (.3282)	3.5736*** (1.1759)	1.6438** (.4774)	1.6208*** (.4077)	1.9837*** (.4404)	1.6892* (.7553)	2.3060* (.8795)	1.73990*** (.3749)
1/Se$_i$ ($\hat{B}_1$)	.0041* (.0019)	0.0032** (.0014)	0.0022 (.0066)	0.0045*** (.0007)	.0014 (.0037)	.0037 (.0026)	.0014 (.0051)	.0044*** (.0010)	.0128^x (.0081)	.0025 (.0019)	.0014 (.0014)	.0049*** (.0004)
N	141	139	94	45	71	70	65	65	107	34	54	85
Studies	37	36	31	8	22	17	15	19	22	15	14	22
R^2	.0358	.0588	.0179	.4748	.0147	.0235	.0140	.1935	.0847	.0616	.0129	.1094
F	4.75	4.79	1.21	43.25	.15	1.99	.94	18.98	2.49	1.66	.90	137.45
Root MSE	4.6386	2.6803	3.159	1.1589	1.7321	6.1223	2.7333	1.8258	4.834	3.46	2.8882	2.5308

a. ***, **, *, ^{x} denotes p values <.001, <.01, <.05 and <.10 respectively; cluster robust errors appear in brackets.

b. If the literature is free of publication bias, the intercept ($\hat{B}_0$) *should not be* statistically significant.

Table 2: MRA (filtered effects - dep. var. corrected t-stats)[a]

	All studies	Without outliers	Cross section	Panel data	Smaller samples	Larger samples	Low HIndex	High HIndex	Robust checks	No checks	Before 2009	After 2009
1/Se$_i$ ($\hat{B}_1$)	.0077** (.0027)	.0061*** (.0018)	0.0053^x (.0028)	0.0070*** (.0002)	.0036 (.0037)	.0068** (.0022)	.0037* (.0017)	.0067*** (.0004)	.0188* (.0091)	.0042* (.0017)	.0053* (.0027)	.0071*** (.0006)
N	141	139	94	45	71	68	65	65	107	34	54	87
Studies	37	36	31	8	22	16	15	19	22	15	14	23
R^2	.1186	.1828	.0976	.5637	.0899	.2118	.0947	.3212	.1962	.1970	.0706	.1821
F	8.16	11.81	0.65	1809.54	0.94	9.54	4.42	231.02	4.24	1.38	3.88	2.25
Root MSE	4.9136	3.011	3.4573	1.6735	1.8600	3.8321	2.9367	2.1347	5.0081	3.5824	3.2518	5.6558

a. ***, **, *, ^{x} denotes p values <.001, <.01, <.05 and <.10 respectively; standard errors appear in brackets.

Table 3: Meta-Significance tests (MST - dep. var. Ln|t|)[a]

	All studies	Without outliers	Cross section	Panel data	Smaller samples	Larger samples	Low HIndex	High HIndex	Robust checks	No checks	Before 2009	After 2009
Intercept ($\hat{B}_0$)	-1.6087* (.7210)	-1.3280* (.6825)	-2.2662* (1.1670)	-1.9044** (.6585)	-3.1157* (1.1210)	1.0878 (1.4352)	.0046 (.7115)	-2.6332*** (.7303)	-1.2164 (.9686)	-1.8416*** (.4328)	-1.6425 (1.1180)	-1.0184^x (.7251)
$Ln\,(df)(\hat{B}_1)$	.2861** (.1046)	.2440* (.0966)	.4009* (.1850)	0.2821** (.0730)	.5368** (.1779)	-.0352 (.1594)	.0577 (.0807)	.4419*** (.1029)	.2229^x (.1276)	.3412*** (.0623)	.2888^x (.1617)	.2019* (.1045)
N	141	139	94	45	71	68	65	65	105	34	54	85
Studies	37	36	31	8	22	16	15	19	21	15	14	22
R^2	.1556	.1183	.1595	.1557	.2084	.001	.0067	.2853	.0834	.3277	.1333	.1034
F	7.48	6.38	4.70	14.92	9.10	0.05	.51	18.45	3.05	29.95	3.19	3.73
Root MSE	1.0471	1.0216	1.0732	.8346	.8851	1.108	.9589	.9746	1.0807	.8059	1.2758	.8340

a. ***, **, *, ^{x} denotes p values <.001, <.01, <.05 and <.10 respectively; cluster robust errors appear in brackets.

Table 4: Nested Multivariate MRAs (FAT-PET, MST)[a, b]

Blocs	Models	A.I FAT t	A.II FAT t	A.III Ln|t|	B.I FAT t	B.II FAT t	B.III Ln|t|	C.I FAT t	C.II FAT t	C.III Ln|t|	D.I FAT t	D.II FAT t	D.III Ln|t|	E.I FAT t	E.II FAT t	E.III Ln|t|
	Intercept ($\hat{B}_0, \hat{b}_0$)	1.899*** (.380)	1.297** (.455)	-1.328^{x} (.683)	1.697*** (.322)	.930* (.405)	-1.515* (.646)	1.800*** (.330)	1.174** (.394)	-1.481* (.669)	1.558*** (.445)	.925* (.420)	-.465 (.563)	1.650*** (.367)	.836* (.387)	.017 (.616)
A. Measures of Precision	*1/Se$_i$* ($\hat{B}_1$)	.003** (.001)			-.003 (.002)			.010 (.059)			.082 (.058)			.110* (.049)		
	$\sqrt{Df}$ ($\hat{B}_1$)		.018* (.008)			.015** (.005)			.011* (.005)			.012^{x} (.006)			.016** (.005)	
	Ln(Df) ($\hat{b}_1$)			.244* (.097)			.329*** (.080)			.292*** (.081)			.208** (.077)			.161* (.067)
B. Data Indicators	Export var. dich. (XVAR)				-.001 (.003)	-.002 (.003)	.005 (.215)	.006 (.004)	.001 (.004)	.181 (.321)	.017 (.016)	.018 (.015)	.164 (.273)			
	Manuf. sector (MANUF)				.033*** (.007)	.030*** (.007)	-.782^{x} (.429)	.016 (.021)	.013 (.023)	-.535 (.386)	-.007 (.023)	-.008 (.029)	-.369 (.406)	.006 (.013)	.007 (.022)	-.237 (.297)
	ASIA data (ASIA)				-.028*** (.006)	-.027*** (.007)	.452 (.397)	-.037 (.069)	-.021 (.025)	.343 (.331)	-.110^{x} (.064)	-.013 (.016)	.183 (.302)	-.132* (.057)	.001 (.022)	.185 (.234)
C. Regressors	Reg. pressures (REGPRESS)							.018 (.015)	.014 (.011)	.192 (.429)	.008 (.007)	.006 (.006)	.310 (.435)			
	Com. pressures (COMPRESS)							-.023*** (.005)	-.021*** (.005)	.108 (.553)	-.120** (.043)	-.090** (.031)	.019 (.469)	-.085** (.037)	-.010 (.010)	.133 (.378)
	Con. pressures (CONPRESS)							.023 (.044)	.027^{x} (.015)	-.252 (.240)	.058* (.032)	.088** (.029)	-.083 (.269)			
	Investor pressures (INVPRESS)							.007 (.025)	.014 (.028)	.522 (.378)	-.013 (.028)	.019 (.026)	.533 (.344)	-.022 (.024)	.025 (.025)	.661** (.204)
	Corp. Soc. Resp. (CSR)							-.012 (.027)	-.007 (.017)	.088 (.410)	.112* (.051)	.092* (.039)	-.118 (.502)	.084* (.040)	.020 (.021)	-.183 (.544)
	High Imp. Ind. (HImpact)							.029 (.040)	.022 (.024)	-.609 (.405)	.075* (.033)	.035* (.015)	-.575 (.380)	.062^{x} (.035)	-.008 (.014)	-.507 (.364)
	Economic perf. (ECONPERF)							.015 (.021)	.011 (.014)	.057 (.279)	.032 (.020)	.010* (.005)	.018 (.306)	.033 (.021)	-.002 (.009)	-.119 (.327)
	Int. networks (IN)							-.031 (.028)	-.023 (.017)	-.014 (.334)	-.013 (.013)	.004 (.009)	.024 (.332)	-.008 (.010)	.015 (.010)	.077 (.270)
D. Quality Indicators	Study pub.>2009 (AFTER2009)										.097* (.044)	.062* (.025)	.024 (.297)	.089* (.042)	.018 (.019)	.036 (.334)
	H-Index (H-INDEX)										-.013^{x} (.007)	-.007^{x} (.004)	-.040 (.047)	-.015* (.007)	-.003 (.003)	-.036 (.036)
	Est. method (PROB/LOGIT)										-.005^{x} (.003)	-.004^{x} (.003)	-.584* (.260)	-.006* (.003)	-.007* (.003)	-.683** (.266)
	R-squared	.059	.066	.118	.258	.297	.158	.369	.391	.284	.553	.512	.336	.509	.392	.324
	Root MSE	2.680	2.671	1.022	2.407	2.342	1.010	2.289	2.249	.960	1.949	2.037	.935	2.018	2.246	.933
Nested Model Tests[c] Criteria: $F \geq F_{.005,\, p,\, n-(k+p+1)}$					$F_{A\to B}$ 11.97	$F_{A\to B}$ 14.74	$F_{A\to B}$ 1.94	$F_{B\to C}$ 2.78	$F_{B\to C}$ 2.78	$F_{B\to C}$ 2.41	$F_{C\to D}$ 16.71	$F_{C\to D}$ 10.21	$F_{C\to D}$ 35.74			

a. ***, **, *, ^{x} denotes p values <.001, <.01, <.05 and <.10 respectively; cluster robust errors appear in brackets.

b. n=139 from a total of 36 studies.

c. k=parameters in the restricted model, p=additional parameters in the full model. The shaded cell indicates a non-significant nested model.

Appendix

Table A1: Studies included in the Meta-Regressions

	Authors	Journal	Pub. Year	No. Obs.	No. Regs	No. Controls
1	Arimura, T. H., Hibiki, A., Katayama, H.	Is a voluntary approach an effective environmental policy instrument? A case for environmental management systems. *Journal of Environmental Economics and Management*, *55*(3), 281-295.	2008	792	2	32
2	Arimura, T. H., Darnall, N., Katayama, H.	Is ISO 14001 a gateway to more advanced voluntary action? *Journal of Environmental Economics and Management*. 61(2), 170-182.	2011	945	2	40
3	Baek, K.	The Diffusion of voluntary environmental programs: The case of ISO 14001 in Korea, 1996–2011. *Journal of Business Ethics*, 145, 325–336.	2017	6631	6	8
4	Blackman, A., Guerrero,S.	What drives voluntary eco-certification in Mexico? *Journal of Comparative Economics*, 40, 256-268	2012	30605	2	15
5	Bluffstone, R., Sterner, T.	Explaining environmental management in Central and Eastern Europe. *Journal of Comparative Economic Studies,* 48, 619-640.	2006	1049	2	48
6	Chapple, W., Cooke, A., Galt, V., Paton, D.	The characteristics and attributes of UK firms obtaining accreditation to ISO 14001. *Business Strategy and the Environment,* 10, 238-244.	2001	13766	7	6.7
7	Christmann, P., Taylor, G.	Globalization and the environment: Determinants of firm self-regulation in China. *Journal of International Business Studies, 32,439-458.*	2001	86	4	16.5
8	Cole, M. A., Elliott, R.J.R., Shimamoto, K.	Globalization, firm-level characteristics and environmental management: A study of Japan. *Ecological Economics*, 59, 312-323.	2006	394	1	12
9	Darnall, N., Kim, Y,	Which types of environmental management systems are related to greater environmental improvements? *Public Administration Review, 73(3), 351-365.*	2012	821.2	10	20
10	Delmas, M., Montiel, I.	Greening the supply chain: When is customer pressure effective? *Journal of Economics and Management Strategy, 18(1), 171-201.*	2009	10383	20	18.3
11	Delmas, M., Pekovic, S.	Environmental standards and labor productivity: Understanding the mechanisms that sustain sustainability, *Journal of Organizational Behavior*, 34,230–252.	2013	4929	2	26
12	Ferron-Vichez, V., Darnall, N.	Two are better than one: the link between management systems and business performance. *Business Strategy and the Environment,* 25(4), 221-240.	2016	2699	3	18
13	Freitas, I.M.B., Izuka, M.	Openness to international markets and the diffusion of standards compliance in Latin America. A multi-level analysis, *Research Policy,* 41:201-215.	2012	43	1	12
14	Grolleau, G., Mzoughi, N., Pekovic, S.	The characteristics of chemical firms registering for ISO 14001 or Responsible Care, *Economic Bulletin, 29(29),1-13.*	2007	86	1	10
15	Grolleau, G., Mzoughi, N., Pekovic, S.	What drives agrifood firms to register for an environmental management systems? *European Review of Agricultural Economics,* 34(2), 233-255.	2007	215	1	12
16	Grolleau, G., Mzoughi, N., Pekovic, S.	Environmental management practices: good or bad news for innovations delivering environmental benefits? The moderating effect of market characteristics.*Economics of Innovation and New Technology*, 24(4), 339-359.	2015	4114	1	20
17	Gumerotti, N. M., Testa, F., Amirante, D. and Frey, M.	The role of negotiating tools in the environmental policy mix instruments: Determinants and effects of environmental agreement, *Instituto di Management, Scuola Superiore Sant' Anna,* WP, n. 02/2012	2012	1712	1	20
18	Henriques, I, Husted, B. W., Montiel, I	Spillover effects of voluntary environmental programs on greenhouse gas emissions: Lessons from Mexico, *Journal of Policy Analysis and Management*, 32(2), 296-322.	2013	1149	1	12

19	Johnstone, N.,Labonne, J.	Why do manufacturing facilities introduce environmental management systems? Improving and/or signaling performance. *Ecological Economics, 68,719-730.*	2009	1928	5	15.2
20	King, A.A., Lenox, M.J. Terlaak, A.	The strategic use of decentralized institutions: Exploring certification with the ISO 14001 Management Standard. *Academy of Management Journal, 48(6):1091-1106.*	2005	19084	3	33
21	Melnyk, S.A, Sroufe, R.P, Calantone, R. J.	A model of site-specific antecedents of ISO 14001 certification. *Production and Operations Management,* 12(3), 369-385.	2003	1451	2	17
22	McGuire, W.	The effect of ISO 14001 on environmental regulatory compliance in China. *Ecological Economics,* 105, 254-264.	2014	580	6	32.5
23	Montiel, I, Husted, B. W.	The adoption of voluntary environmental management programs in Mexico: First movers as institutional entrepreneurs. *Journal of Business Ethics,* 88, 349-363.	2009	1328	1	14
24	Montiel, I., Husted, B. W., Christmann, P.	Using private management standard certification to reduce information asymmetries in corrupt environments*, Strategic Management Journal, 33, 1103-1113.*	2012	433	2	7
25	Mori, Y., Welch, E. W.	The ISO14001 environmental management standard in Japan: results from a national survey of facilities in four industries, *Journal of Environmental Planning and Management,* 51(3):421-445	2008	1484	3	29
26	Nakamura, M., Takahashi, T.,Vertinsky, I.	Why Japanese firms choose to certify: A study of managerial responses to environmental issues, *Journal of Environmental Economics and Management,* 42, 23-52.	2001	193	10	34.1
27	Nishitani, K	An empirical study of the initial adoption of ISO 14001 in Japanese manufacturing firms*. Ecological Economics,* 68, 669-679.	2009	741	4	27
28	Qi, G., Zeng, S., Yin, H., Lin, H.	ISO and OHSAS certifications: How stakeholders affect corporate decisions on sustainability. *Management Decision,* 51, 1983-2005.	2013	903	1	11
29	Ruiz-Tagle, M. T.	Why do manufacturing plants invest in environmental management? WP *No. 202006, University of Cambridge*	2006	4222	7	25.4
30	Shen, J.Y., Qin, X-D.	What determines Chinese firms' decision on implementing voluntary environmental schemes? *Journal of Service Science and Management,* 4, 380-390.	2011	270	2	16
31	Takahashi, T.,Nakamura, M.	The impact of operational characteristics on firms' EMS decisions: Strategic adoption of ISO14001 certifications, *Corporate Social Responsibility & Environmental Management,* 17, 215-229.	2010	879	6	14.5
32	Tambunlertchai, K., Kontoleon, A., Khanna, M.	Assessing participation in voluntary environmental programmes in the developing world: The role of FDI and export orientation on ISO 14001 adoption in Thailand. *Applied Economics,* 45(15), 2039-2048.	2013	494	6	19
33	Uchida, T.,Ferraro, P.J.	Voluntary development of environmental management systems: Motivations and regulatory implications. *Journal of Regulatory Economics,* 32, 37-65.	2007	1154	6	24.2
34	Ullah, B. & Wei, Z	ISO certification, corruption, and firm performance: A cross-country study, *Working Paper, World Bank.*	2014	41553	2	11
35	Wu, S.Y, Chu, P-Y, Liu, T.Y.	Determinants of a firm's ISO14001 certification: An empirical study of Taiwan. *Pacific Economic Review,* 12(4), 467-487.	2007	6187	5	31
36	Zhu,O, Cordeiro,J., Sarkis, J.	International and domestic pressures and responses of Chinese firms to greening*. Ecological Economics,* 83, 144-153.	2012	377	1	14
37	Zhu,O, Cordeiro,J., Sarkis, J.	Institutional pressures, dynamic capabilities and environmental management systems: Investigating the ISO9000-Environmental management system implementation linkage*. Journal of Environmental Management,* 114, 232-242.	2013	377	2	12.5
	Total- 37 studies; 141 regressions	*Average per study*	*2009*	*4435*	*3.8*	*19.8*

Table A2: Independent variables used in ISO14001 adoption studies[b]

Regression accounts for	Frequency	Percent	Regression accounts for	Frequency	Percent
firm size	138	97.9%	exports to China	14	9.9%
industry sector	128	90.8%	exports to Europe	14	9.9%
regulatory pressures	83	58.9%	stakeholder pressures (generalised)	14	9.9%
ISO 9000 or other QMS	83	58.9%	importance of corporate image	14	9.9%
consumer pressures	59	41.8%	disclosure requirements	14	9.9%
foreign ownership	54	38.3%	industry association pressures[c]	14	9.9%
economic performance	49	34.8%	financial institutions pressures	14	9.9%
environmental performance	46	32.6%	CSR status	13	9.4%
community pressures	42	29.8%	quality of EMS	12	8.5%
industry impact on the environment[a]	41	29.1%	labour unions	12	8.5%
NGO pressures	41	29.1%	manager pressures/attitudes	11	7.8%
foreign HQ	38	27.0%	science park location	11	7.8%
publicly traded firms	38	27.0%	Keiretsu, Maquila, or other.	11	7.8%
industrial concentration	37	26.2%	firm supplies government contracts	9	6.4%
R&D expenditures	29	20.6%	FDIs	9	6.4%
firm's age	29	20.6%	exports to developed countries	8	5.7%
exports to OECD countries	27	19.1%	age of firm's assets	8	5.7%
firm's investment in advertising	25	17.7%	firm's permits	7	5%
government subsidies	23	16.3%	MNE ownership	7	5%
firm's location within a country	17	12.1%	firm's exposure to risk	6	4.3%
firm's employees age	17	12.1%	state ownership	6	4.3%
ability to compete on quality	15	10.6%	employee pressures/attitudes	4	2.8%
political pressures	15	10.6%	firm imports	3	2.1%
inspections	15	10.6%	firm number of fines	3	2.1%
firm's capital intensity	15	10.6%	exports to North America	2	1.4%

[a] A firm's environmental impact differs from its environmental performance. A firm can be operating in a highly polluting sector (*i.e.* high impact) but perform well with respect to environmental performance expectations of its industry.
[b] Only variables used in at least 2% of regressions are included.
[c] A few studies explicitly modelled industry associations. These differ from industry membership.

Table A3: Descriptive statistics (n=141)

	All Studies		Without Outlier	
Dependent and precision-related variables - *Bloc A*	**Mean**	**s.d.**	**Mean**	**s.d.**
t-statistic of export coefficient (*t*)	2.675	4.703	2.223	2.750
Logarithm of absolute value of t-statistic (*Ln(|t|)*)	.515	1.135	.4719	1.084
Export coefficient (α)	.577	2.272	.5831	2.288
Partial correlation coefficient (*r*)	.052	.065	.050	.064
Standard error of export coefficient (*SE*)	.289	.845	.293	.850
Degrees of Freedom (*Df*)	4906.333	8491.301	4379.043	7307.78
Independent variables (data controls) – *Bloc B*				
1 if the export variable is dichotomous (*XVAR*)	.433	.497	.425	.496
1 if the data is cross-sectional (*CROSS*)	.667	.473	.676	.470
1 if the data only includes firms from the manufacturing sector (*MANUF*)	.908	.290	.921	.271
1 if ASIA data is used (*ASIA*)	.610	.490	.619	.487
Independent variables (regression explanatory controls) – *Bloc C*				
1 if the regression includes regulatory pressures (*REGPRESS*)	.603	.491	.597	.492
1 if the regression includes community pressures (*COMPRESS*)	.312	.465	.317	.467
1 if the regression includes consumer pressures (*CONPRESS*)	.418	.495	.425	.496
1 if the regression includes investor pressures (*INVPRESS*)	.262	.442	.266	.444
1 if the regression accounts for a firm's corporate social responsibility (*CSR*)	.092	.290	.094	.292
1 if the regression accounts for industry environmental impact (*HIGHIMPACT*)	.319	.468	.326	.470
1 if the regression accounts for the firm's economic performance (*ECONPERF*)	.340	.476	.345	.477
1 if the regression accounts for other forms of international networks (*IN*)	.731	.495	.727	.447
Independent variables (study quality controls) – *Bloc D*				
1 if study was published after 2009 (*AFTER2009*)	.617	.488	.612	.489
1 if the regression is a simple probit or logit (*PROB/LOGIT*)	.681	.468	.691	.464
H-Index of journal (*HINDEX*)	95.90	51.71	97.42	52.84

Table A4: MRA tests for publication selection (dep. var. t-stats)[a,b]

	All studies	Without outliers	Cross section	Panel data	Smaller samples	Larger samples	Low HIndex	High HIndex	Robust checks	No checks	Before 2009	After 2009
Intercept ($\hat{B}_0$)	-.0738 (1.2708)	-1.2968*** (4.552)	-.3098 (1.0440)	1.6806* (.6528)	-.2431 (.8205)	3.6039* (1.4507)	.8020 (.9835)	.9232 (.5858)	1.4889* (.6559)	.9530 (.6809)	1.7133* (.9158)	1.0037* (.5037)
$\sqrt{df}(\hat{B}_1)$	.0501 (.0311)	0.1757* (.0077)	0.0253 (.0132)	0.0072 (.0043)	.0660 (.0441)	-.0045 (.0113)	.0145 (.0082)	.0305 (.0185)	.0129 (0.0084)	.0310 (.0177)	.0163 (.0130)	.0189ˣ (.0096)
N	141	139	94	45	71	68	65	65	105	34	54	85
Studies	37	36	31	8	22	16	15	15	21	15	14	22
R^2	.2165	.0656	.2378	.0362	.0963	.0029	.0508	.2299	.0403	.1459	.0566	.0766
F	2.60	5.23	3.59	2.81	2.24	.18	3.12	2.71	2.33	3.05	1.57	3.85
Root MSE	4.1778	2.6705	2.7829	1.5698	1.6588	3.2695	.2.6818	1.7841	2.4369	3.3008	2.8235	2.577

a. ***, **, *, ˣ denotes p values <.001, <.01, <.05 and <.10 respectively; cluster robust errors appear in brackets.
b. If the literature is free of publication bias, the intercept ($\hat{B}_0$) should not be statistically significant.

Table A5: MRA (FAIVEHR) tests for genuine effect (dep. var. t-stats)[a]

	2-Stage Least Square (instrument $\sqrt{n}$ for $1/Se_i$)											
	All studies	Without outliers	Cross section	Panel data	Smaller samples	Larger samples	Low HIndex	High HIndex	Robust checks	No checks	Before 2009	After 2009
Intercept ($\hat{B}_0$)	-22.8845 (71.1302)	-13.5361 (97.9694)	.1374 (.1775)	3.4253* (1.5166)	.2778 (.8729)	2.7683*** (.7110)	2.8381*** (.7493)	1.02544* (.6117)	4.2518 (3.1672)	-.7446 (2.0441)	5.9491 (5.4887)	.8277 (.7224)
$1/Se_i$ ($\hat{B}_1$)	.2460 (.7152)	.1527 (.9793)	-10.1884 (15.7477)	-.0084 (.0091)	.0144 (.0134)	.0032 (.0072)	-.0087 (.0087)	.0113 (.0072)	-.0329 (.0482)	.0131 (.0094)	-.0194 (.0313)	.0212 (.0197)
N	141	139	94	45	71	68	65	65	105	34	54	85
Studies	37	36	31	8	22	16	15	15	21	15	14	22
Wald Chi2(1)	.12	.02	.60	.85	1.16	.19	1.00	2.46	.47	1.95	.38	1.16
Prob>Chi2	.7310	.8753	.4386	.3568	.2814	.6604	.3177	.1169	.4939	.1629	.5365	.2819
Root MSE	50.762	31.759	26.508	3.2759	2.5414	3.1425	3.5375	2.2561	5.0773	4.9917	5.6485	3.8183

a. ***, **, *, ˣ denotes p values <.001, <.01, <.05 and <.10 respectively; cluster robust errors appear in brackets.